

New fragment weighting scheme for the bayesian inference network in ligand-based virtual screening

Abstract:

Many of the conventional similarity methods assume that molecular fragments that do not relate to biological activity carry the same weight as the important ones. One possible approach to this problem is to use the Bayesian inference network (BIN), which models molecules and reference structures as probabilistic inference networks. The relationships between molecules and reference structures in the Bayesian network are encoded using a set of conditional probability distributions, which can be estimated by the fragment weighting function, a function of the frequencies of the fragments in the molecule or the reference structure as well as throughout the collection. The weighting function combines one or more fragment weighting schemes. In this paper, we have investigated five different weighting functions and present a new fragment weighting scheme. Later on, these functions were modified to combine the new weighting scheme. Simulated virtual screening experiments with the MDL Drug Data Report²³ and maximum unbiased validation data sets show that the use of new weighting scheme can provide significantly more effective screening when compared with the use of current weighting schemes.